图书情报工作
第 62 卷 第 2 期 2018 年 1 月

搜索前后用户知识水平的评估及其变化情况分析

■ 宋筱璇 刘畅
北京大学信息管理系 北京 100871

摘要：[**目的/意义**]信息搜索是人们常用的信息查询方法，目前搜索系统在查找事实型信息时支持效果较好，但是对人们以学习为目的的搜索功能还缺乏研究。“搜索即学习”(search as learning)是近年来交互式信息检索的研究热点，这类研究中将搜索看作学习的过程，并尝试对用户搜索中的知识学习进行评估，进而提出系统支持用户学习的功能优化建议。本文着重解决如何全面评估用户搜索前后的知识水平，为此类研究提供参考。[**方法/过程**]采用用户实验法，对用户搜索前和搜索后撰写的知识内容进行评估，提出综合数量与质量维度的用户知识评估方法，对用户在学习型任务搜索前后知识水平进行评估。数据分析阶段采用统计方法来验证用户搜索后与搜索前的知识水平差异。[**结果/结论**]研究发现，用户在知识数量上的表现随着搜索的完成而变得更加全面和深入，在知识点数量、知识面数量、知识面广度和知识面深度上都有显著的提升。同时，在搜索后产生了专业度较高的知识面。对于搜索前较模糊的某些概念，在搜索后表达得更清晰明确。在质量上，搜索后绝大多数的用户都在知识的相关性、分析程度及用户观点的提出方面有所提升。

关键词：搜索即学习 知识评估 知识变化
分类号：G251
DOI：10.13266/j.issn.0252－3116.2018.02.015

1 引言

随着信息技术和网络技术的快速发展，网络上的信息非常丰富，互联网用户已习惯通过使用网络搜索系统查询和获取各类信息。目前的搜索引擎在查找事实型信息的需求满足上已经做得很好，但是当用户面临探索性检索任务或理解型、分析型学习任务时，搜索系统应该如何有效地提供信息并帮助用户完成学习任务仍有许多有待解决的问题[1]。近年来，搜索即学习(search as learning)是当前交互式信息检索领域研究的热点之一[1-3]。G. Marchionini 曾提出搜索作为一个信息发现的过程，用户不是简单地与系统进行信息交互，而是在过程中也不断地学习和理解，产生了超越信息任务之外的知识增长[4]。用户在搜索过程中的知识变化对用户检索过程的信息行为、交互表现和搜索的效果等方面都可能产生影响，进而影响信息搜索的下一步策略，甚至使整个搜索结果发生改变。对于以学习为目的的搜索来说，用户在知识上的变化或搜索学习产出是评估搜索系统效能的标准之一。近年来研究者们开始从理论上探讨信息搜索过程中的知识学习问题，并尝试提出多种评估方法对用户搜索中的知识学习水平进行评估[3,5]。对于这些评估方法，研究者仍在探索过程中，提出了用户知识的数量、表述知识的长度等一些较为直观的维度的测量，也曾经尝试用概念图的方式衡量搜索对于用户知识结构的改变。本研究不仅关注搜索给用户带来的知识点或节点的数量的变化，也关注用户产生的知识的理解程度、分析程度等质量的变化，并基于此提出多维度知识水平测量的方法，通过实验法来测量用户搜索前后知识水平及其变化情况。实验过程中被试搜索 4 个与学习相关的搜索任务，并在搜索前后撰写与搜索任务相关的文本答案。本文对用户搜索前后的文本进行分析，比较用户搜索前后对于同一个搜索任务的知识水平的变化，具体包括知识数量(知识点数、知识面数、知识面广度、知识面深度)和知识质量(知识相关性、用户观点、知识分析程度)两个维度，并依据评估结果比较用户搜索前后相应知识维度的差异。

作者简介：宋筱璇(ORCID：0000－0002－6589－4071)，博士研究生；刘畅(ORCID：0000－0002－9183－6385)，副教授，博士，硕士生导师，通讯作者，E-mail：imliuc@pku.edu.cn。
收稿日期：2017－08－11 修回日期：2017－11－20 本文起止页码：108－116 本文责任编辑：刘远颖

2 相关研究

2.1 探索式搜索与知识学习

近年来,“搜索即学习”成为交互式信息检索的研究热点,如 2012 年 SWIRL(strategic workshop on information retrieval in Lorne)的研讨会[6]、2013 年的 Dagstuhl 高峰论坛[7]、2014 年美国信息科学与技术学会(ASIS&T)第 77 届年会[3]、2016 年国际计算机协会信息检索国际会议(SIGIR)[2]等重要的国际会议上都有关于这个主题的专项研讨。关于搜索即学习,学者提出关注搜索过程中用户的学习过程,搜索任务及其他搜索情境对用户学习的影响,关注学习效果的测量与评估方法以及如何优化系统以便支持用户在搜索过程中的学习。

关于用户搜索中知识学习的研究是与探索式搜索任务的相关研究紧密联系的。G. Marchionini[4]提出了探索式搜索,认为探索式搜索区分于事实查找的搜索任务,除了搜索基本的事实和信息,更侧重于知识的获取,概念的理解,观点的阐释、对比或数据概念的聚合等几方面。探索式搜索通常是指在不熟悉目标领域、不清楚搜索目标、不确定达到目标的路径的情况下,用户不能直接从搜索系统中获得所需答案,只能通过系统帮助不断探索、尝试,进而消除不确定性[8]。探索式搜索的相关研究表明在探索式搜索下用户交互行为的特征和表现与事实型搜索有较大差异。例如袁红和李秋[9]发现对于提问应答式搜索,用户倾向于将问题本身作为检索词;而在探索式搜索中,任务越复杂,用户越倾向使用检索词扩展进行搜索,且检索词数量越多,词频越低,词间的联系更不紧密。同时,探索式搜索中浏览网页的数量也比提问应答式更多,链接深度更深。张云秋等[5]发现在探索式搜索中,随着搜索主题难度的增加,页面的平均浏览时间减少,更难集中在某一网页上进行精读。张云秋等[10]也尝试对探索式搜索过程中用户知识结构的改变进行了分析,通过被试在搜索前、后绘制的任务知识概念图,发现在探索式搜索任务中,搜索者的知识结构会有较大改变,搜索者在搜索过程中试图构建新的、复杂的和更为专业的知识结构。探索式搜索包含的范围很广,学习相关的搜索是探索式搜索中的一种,是以学习为目的的搜索,需要对搜索信息的“how”和“why”有更加深入的理解[11]。K. Collins-thompson 等提出我们应关注以学习为导向的探索式搜索(learning-oriented exploratory search)[12],学习是搜索中的重要认知过程,学习效果也是评价探索式搜索效果的重要指标之一[8]。

2.2 用户知识水平评估

用户搜索前已有的知识水平是信息检索领域的重要用户特征,许多研究都证实了用户搜索前的知识水平会影响其信息行为及交互特征[13－15]。用户搜索后的知识水平被认为是评价用户搜索效果的重要指标之一[8]。因此,知识水平的评估是研究用户搜索中,特别是学习型搜索中的核心问题。

在信息搜索行为研究中,学者曾尝试提出过若干种评估用户知识水平的方式。如通过量表的方式让用户对自身的知识水平进行自我评价[13－14,16],这种评估方式相对主观,在很大程度上依靠问卷的有效性或是个人评价的准确情况。但事实证明用户对于自身知识水平的评价未必准确,存在用户高估或低估个人水平的情况[17]。

部分研究者通过测试的方式,让用户在搜索前或搜索后作答与任务相关的问题,这类问题往往具有明确且唯一的答案,测试的得分即代表用户当前的知识水平。例如 W. R. Hersh 等[18]让医学院的学生在搜索前后分别对一些医学相关且答案明确的问题进行回答,并对他们回答的正误进行打分。通过学生前后得分的差异,说明信息搜索对于学生知识提升的作用。L. Nelson 等[19]在探索大众标注对搜索学习的影响作用时,让用户对“企业 2.0 混合”这一任务进行时长两小时的搜索。在搜索后,回答实验者设计的该领域的 20 个判断题,以得分表示用户进行搜索后的学习效果。

另外,也有研究考量的是用户搜索后完成的任务摘要的质量,以此来对知识水平进行评估。任务摘要指的是用户根据任务设计的要求,完成与任务相关的小段文字报告或文字回答。例如 Y. Kammer 等[20]根据用户任务完成过程中撰写的摘要和提供的关键词对用户搜索后的知识状态进行评估。其中在关键词方面,统计有价值的关键词的数量;而在任务摘要上,主要计算用户找到的论据的数量以及罗列的相关组织或个人的数量。T. Willoughby 等[21]探索学生的领域知识对其学习效果的影响时,要求学生针对任务主题,在搜索后完成一篇任务短文,通过计算短文中可接受且正确的陈述或短语的数量,对学生完成的短文进行评分。

以上研究均侧重知识点的数量的测量,M. J. Wilson 和 M. L. Wilson[22]提出知识质量也是评估用户知识水平的很重要的维度。他们依据 B. S. Bloom 等[23]

的认知学习领域分类法，建立了学习深度测量(depth of learning)的方法，对用户信息检索前后所写的任务摘要进行了有用性(该信息是否有用，4 点量表)、分析性(该信息用户是否进行了分析处理，3 点量表)、评价性(该信息中是否包含用户的个人评价，2 点量表)的人工评判，以表示搜索是否为用户带来了知识变化。此外，该研究还对比了传统的任务摘要处理办法，计算事实(facts)和陈述(statements)的数量以及摘要中话题的覆盖情况等，结果发现单一计算事实和陈述的数量将导致知识增长情况在很大程度上受到摘要长度的影响，而且仅对事实和陈述的数量、话题覆盖的数量进行计算很难表现出该信息的质量。因此，M. J. Wilson 和 M. L. Wilson[22] 认为对于知识的测量不应该片面地使用诸如事实项、分面覆盖度等某一种测量方式，而应该多方面测量，综合表示用户当前的知识状态。本文综合上述研究所提出的评估方式，对用户搜索前后知识采用数量(知识点、知识面、知识面广度、知识面深度)和质量(相关性、用户观点、分析程度)相结合的评估体系，力图更全面综合地展现用户的知识水平。

3 研究方法

本研究主要采用实验法进行研究，本节将具体介绍用户实验的被试情况、实验流程设计、实验任务设计的具体细节。在用户知识水平评估部分，阐释本文建立的多维度的知识水平评估体系的具体内容，并详细描述基于该体系的数据编码过程及结果。

3.1 用户实验

本实验邀请了来自北京大学的 32 名实验参与者。男性 12 人，女性 20 人，年龄分布在 18 – 30 岁之间。他们来自不同的院系：人文社科类 14 人、信息与工程科学类 5 人、理学类 10 人、医学类 3 人，且分布于不同的年级：本科生 20 人(大一 8 人、大二 3 人、大三 4 人、大四 5 人)、硕士研究生 4 人、博士研究生 8 人。实验过程中，实验参与者不受任何限制或干扰，以使其尽量自然地模拟他们在现实生活中的搜索行为。

3.1.1　实验流程设计　实验开始前，参与者首先完成一个基本情况调研的问卷和认知类型的测试，然后开始对 4 个搜索任务逐一进行搜索。在每个任务搜索前，实验参与者需完成搜索前问卷，对该任务的熟悉度、任务难度进行量表评价，同时用文字描述对该任务已经了解的信息和可能选用的检索词。接着实验参与者正式对任务进行搜索，时间不限，在搜索过程中记录和整理自己认为合理的答案，并将文本记录在电脑的记事本文件中。当被试认为自己的记事本文件中已完整记录了该搜索任务的答案后，即可随时停止搜索。搜索结束后，实验参与者关闭浏览器及搜索过程中记录的记事本文件，回答搜索后问卷，对任务完成的满意度、搜索之后任务的熟悉度、网络信息的丰富度、任务的难度进行量表评价，并在不参考外在信息(如相关网页和记事本文件)的情况下，通过记忆，用文字描述对该搜索任务的回答。4 个搜索任务全部完成之后，实验参与者填写实验后问卷，对其搜索体验和搜索过程进行评价。

实验中利用 Morae Recorder 3.3 在后台记录用户搜索交互行为，包括点击页面的 URL，每个页面上停留时间、阅读时间、该页面上鼠标和键盘活动情况、该页面的浏览次数，不同页面上的相对活动情况，用户检索式的输入和修改，用户检索式与页面活动情况关系等方面。

3.1.2　实验任务设计　本实验的目的是对用户搜索前后知识的变化特征进行探索。因此，在实验任务的设计上，围绕用户知识学习这一中心，以 J. Lee[24] 的认知学习模式作为任务设计的理论依据。J. Lee[24] 在 B. S. Bloom 等[23] 认知学习分类法的基础上，提出了接受型学习(receptive learning)、评价型学习(critical learning)和创造型学习(creative learning)3 类新的认知学习模式。其中接受型学习指的是学生对所学习的知识的记忆、理解和表达，对应 B. S. Bloom 分类中的识记(remembering)和理解(understanding)层面，是一种相对浅层的学习。评价型学习指的是通过对不同来源的知识进行反思、批判和评估，从而形成学生自身的思考的过程，对应 B. S. Bloom 分类中的应用(applying)、分析(analyzing)和评价(evaluating)层面，是一种比接受型学习更为深层的学习过程。创造型学习对应 B. S. Bloom 分类中的创造(creating)层面，是一种要求最高的学习模式，需要学生对所学习的知识生成新的想法或构建出新的结构。

本实验选择了大学生较关注的日常生活健康领域相关问题进行任务主题的设计，基于接受型和评价型两个学习层面，设计了接受型学习任务和评价型学习任务两类不同的任务。每名实验参与者需要对两类学习任务展开搜索，每类搜索任务包括两个不同的搜索主题，因此每个被试共搜索 4 个任务，任务顺序采用拉丁方轮换方法，尽可能减少题目顺序和疲惫感给被试带来的影响[25]。不同类型的任务描述示例如下：

(1)接受型学习任务：北京的雾霾越来越严重，你

觉得非常有必要在雾霾天气时佩戴口罩。在买口罩前你想知道口罩都有哪些种类,不同种类之间的口罩区别是什么,哪种口罩适合年轻人在北京长期佩戴。请在网络上搜索,并在搜索的过程中在记事本中记录下你的答案。

(2)评价型学习任务:你上初中的表弟正在考虑加入学校的足球队,大部分亲戚均对这个主意感到赞同。但也有人提出足球是一项危险的运动,并且担心会有潜在的健康危险。他的父母也想问问你的意见:长期踢足球对青少年的健康会有哪些危害?是否应该让孩子加入足球队。请在网络上搜索,并在搜索的过程中在记事本中记录下你的答案。

3.2 用户知识水平评估

本研究以用户搜索前后对搜索任务问题回答的文字为基础,从回答文字的定量和定性两个维度综合评估被试的知识水平。定量的指标包括知识点、知识面、知识面广度、知识面深度,定性的评估指标包括知识相关性、用户观点、知识分析程度等。具体评估指标及解释如表1所示:

表1 搜索用户知识评估指标及其解释

知识评估变量	变量说明
知识点数	在用户的回答中有效的名词及数量词个数
知识面数	在用户的回答中涉及的与搜索问题相关的主题数
知识面广度	该回答提及的知识面数/该任务中的知识面总数
知识面深度	该任务中用户提及的知识点数/用户提及的知识面数
知识相关性	0分:回答了和任务完全不相关或是没有提供什么有用的信息
	1分:回答了相对宽泛的、有一点用的信息
	2分:对于完成任务提供了很有用的信息
用户观点	0分:没有提出用户个人的看法或观点
	1分:提出了用户个人的看法或观点
知识分析程度	0分:未对观点或问题进行深入分析(区别、利弊、优劣等的讨论)
	1分:对于观点或任务有初步简单的分析(区别、利弊、优劣等的讨论)
	2分:对于观点或任务有较为详尽的分析(区别、利弊、优劣等的讨论)

为了获得以上变量的信息,本研究采用了三步分析法:

第一步:分词及词性划分。首先利用北京理工大学张华平博士的 NLPIR 汉语分词系统(又名 ICTCLAS2013)对用户搜索前后的回答进行分词和词性的划分。

第二步:知识点和知识面的人工编码。分词完成之后选择对其中的名词和数量词进行知识点的人工编码。编码主要包括两个方面的内容:一是该词是否能被接受作为知识点;二是该知识点属于任务的哪个知识面。人工标注过程由两名编码人员完成,当两人意见不一致时,由独立的第三位编码人员对不一致数据进行再次编码。任务的知识面在编码的过程中不断归纳调整,最终,对两名编码人员知识点的编码一致性进行检验,口罩任务中 Kappa = 0.841,足球任务中 Kappa = 0.883,均大于0.75,信度较高。

根据编码结果,接受型学习任务(口罩任务)形成了8个知识面,分别是型号、材料、功能、外观、性质、认可、使用和价格,表2列出了与该搜索任务相关的8个知识面的若干个知识点。

表2 接受型学习(口罩)任务知识面及知识点编码

口罩知识面	所属知识点示例
型号	N95、N98、KN90
材料	化纤、活性炭、棉质
功能	防霾、防菌、防寒
外观	耳挂式、头戴式、罩杯式
性质	透气性、密闭性、过滤性
认可	品牌、标准、评价
使用	使用次数、寿命、周期
价格	元、20-30

评价型学习任务(足球任务)形成了12个知识面,分别是骨骼肌肉、大脑、内脏器官、软组织、系统调节、其他身体损伤、身体机能锻炼、学业、性格、交际、兴趣和运动防护,表3列出了与该搜索任务相关的12个知识面的若干个知识点。

表3 评价型学习(足球)任务知识面及知识点编码

足球知识面	所属知识点示例
骨骼肌肉	肌肉拉伤、韧带、关节、半月板、骨折
大脑	脑震荡、脑膜炎、脑脊液、头球
内脏器官	内脏、心脏、心律失常
软组织	眼睛、鼻子、颈部软组织
系统调节	呼吸系统、神经系统
其他身体损伤	碰撞、受伤、对抗
身体机能锻炼	身体锻炼、体质、身体发育
学业	学业、学习、智力
性格	意志力、自信、性格
交际	人际交往、团队、合作
兴趣	兴趣、爱好
运动防护	护具、按摩、装备

第三步:知识的相关性、用户观点和分析程度的编码。与用户的知识点和知识面的编码过程类似,两名编码人员分别根据表1,从知识相关性、用户观点和知识分析程度对32名实验者两个任务搜索前后的共128

条回答进行评估。得分不一致的回答由第三位编码人员进行独立的判断。最终知识相关性 Kappa = 0.671,用户是否有观点 Kappa = 0.95,问题分析程度 Kappa = 0.735。

4 数据分析结果

本节通过统计分析比较用户搜索前后知识状态的差异,分别从知识数量(知识点数、知识面数、知识面广度、知识面深度)和知识质量(知识相关性、用户观点、知识分析程度)两个维度展开分析。其中由于知识的数量是数值型变量且满足正态和方差齐性的要求,采用配对样本 t 检验观测搜索前后用户知识状态是否存在差异;知识的质量是定序变量,则采用 Wilcoxon 配对样本检验方法。共计 32 名被试参与到本文的实验中,由于本文对于知识的评估和变化分析属于尝试性研究,因此在本文分析中暂时选择用户的接受型学习任务中的口罩任务和评价型学习任务中的足球话题两个任务展开分析。

4.1 搜索前后用户知识数量的变化

用户搜索前后的知识水平在知识数量上存在显著的差异,如表 4 所示,通过信息搜索,用户的知识点数、知识面数、知识面广度、知识面深度都有显著增加。从各项均值可知,搜索前每个用户在一个任务中平均表达 3-4 个知识点,覆盖 1-2 个知识面,对于每个知识面用 1-2 个知识点来表述。而在完成搜索之后,用户平均表达的知识点超过 11 个,覆盖 3-4 个知识面,对于每个知识面用 3 个左右的知识点来表述。笔者随后针对两种搜索任务分别进行了验证,结果发现无论是哪种搜索任务,用户在搜索后的知识数量都比搜索前有显著提升。

表 4 用户搜索前后知识数量差异 t 检验

知识数量	搜索前(均值(方差))	搜索后(均值(方差))	比较
知识点数	3.75(10.1)	11.16(47.72)	$p<0.0001$ *
知识面数	1.92(1.72)	3.75(3.11)	$p<0.0001$ *
知识面广度	0.21(0.025)	0.40(0.047)	$p<0.0001$ *
知识面深度	1.66(1.35)	2.85(2.22)	$p<0.0001$ *

注:* $p<0.05$

以下具体以用户为单位,从每名用户搜索前后在知识点及知识面上的表现展开分析。如图 1 - 图 4(横、纵坐标分别代表用户在搜索前后的知识点、知识面、知识面广度及深度 4 个指标上的数量分布情况)所示,在 64 个会话中,55 个会话中的知识点数量,50 个会话中的知识面数量、知识面广度,48 个会话中的知识面深度方面的表现都是优于搜索前的。在搜索开始之前,有 10 个会话中的用户感到对任务无话可说,即知识点和知识面数量都为 0,呈现出知识表达上的困难,而这种现象随着搜索的完成,获得了有效的改善,减少至 2 个会话,这个结果说明对于绝大多数用户而言,搜索后的知识水平在知识点与知识面上有所提升。但从图中可以发现,在部分会话中,用户搜索前后的知识点与知识面数量呈现不变,甚至搜索后减少的现象。通过对这部分用户的个案分析发现,对于一部分用户而言,他们的知识点与知识面虽然在数量上没有增加,但具体内容发生了变化。比如有的用户搜索前后都提到了 3 个知识面,但搜索后其实产生了新的知识面,去掉了搜索前的相关度较低的知识面。因此通过搜索,该用户仍然获得了新的知识,这解释了部分用户的知识数量绝对值在搜索后未见提升的原因。

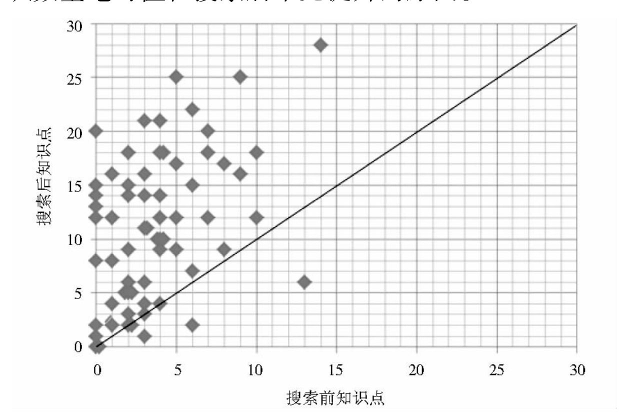


图 1 用户搜索前 - 搜索后知识点数量

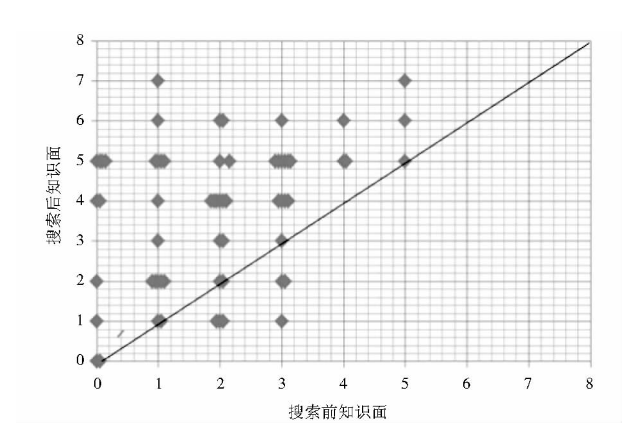


图 2 用户搜索前 - 搜索后知识面数量

从用户角度分析可知,搜索后用户的知识水平在知识点和知识面的表现上较搜索前发生了显著的提升。下面从知识的角度,具体分析搜索前后知识面的覆盖和变化情况,即用户知识点、知识面的提升具体表现在哪些知识上。

在接受型任务(口罩任务)中,如图 5 和图 6 所示,

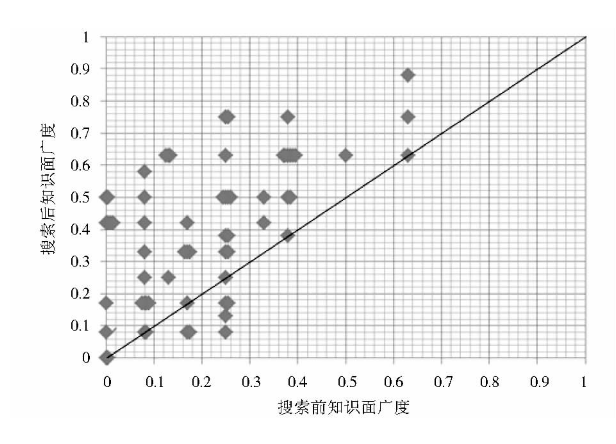


图 3　用户搜索前－搜索后知识面广度

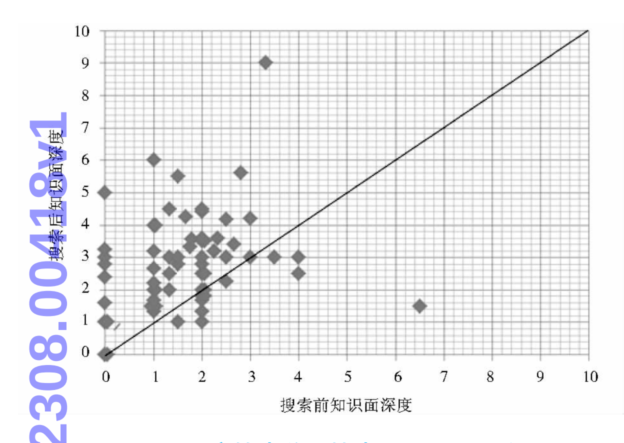


图 4　用户搜索前－搜索后知识面深度

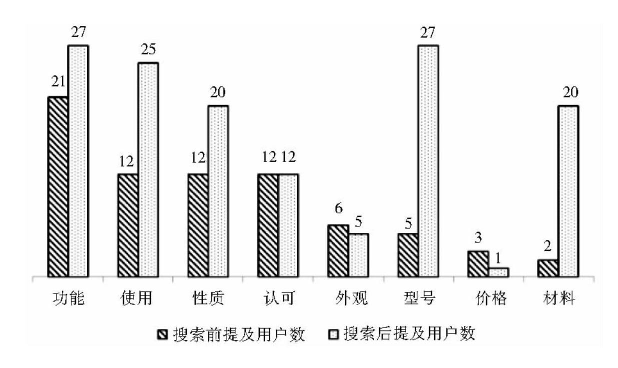


图 5　知识面的用户数量分布(单位:人)

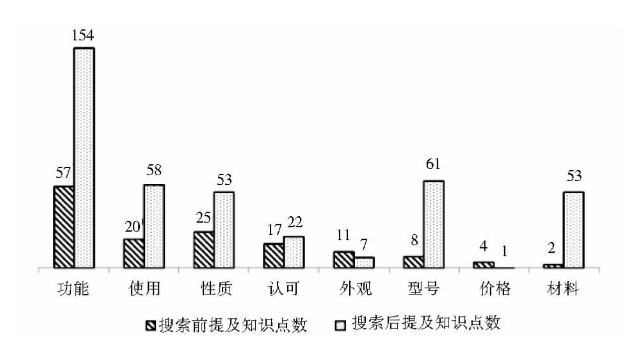


图 6　知识面包含知识点数量分布(单位:个)

用户搜索前后整体提及的知识面没有变化,但在 8 类知识面的分布比例上有较大差异。在搜索前,用户提及最多的是口罩功能方面的问题,有 21 人,共计 57 个知识点,其中防霾,防尘等功能被提及的频次最多。其次是使用、性质和认可方面的内容,都有 12 人提到,20 个左右知识点。而对于型号、材料这类相对而言专业性较强的知识面,则提及人数较少,分别只有 5 人和 2 人。随着用户搜索的深入,情况发生了显著的改变,更多的用户意识到,不同口罩之间的区别与型号和材质有着很大的关系。因此,虽然口罩的功能仍然在其中发挥着巨大的作用(有 27 人涉及),但型号和材料分别上升为 27 人和 20 人提及用户,此外对于性质和使用的关注也有显著提升。

在评价型任务(足球)中,用户搜索前后整体提及的知识面变化则比较明显,如图 7－图 8 所示,搜索后用户产生了运动防护、大脑、性格和系统调节知识面的相关内容,而这是在搜索前用户所没有涉及的。特别是运动防护这一知识面,搜索之前没有用户关注到这方面的问题,但在搜索之后,有近三分之一(10 人,38 个知识点)的用户认为只要做好防护方面的工作,应该主张孩子加入足球队。在搜索前,有 15 人提到骨骼肌肉损伤,18 人提到概念性模糊的损伤,例如对抗、受伤这类知识点。而在用户搜索之后,模糊性损伤的指向性更加明确,80% 左右的用户(25 人)都能很明确地指出足球运动对骨骼肌肉上的伤害,且与骨骼肌肉知识面相关的知识点由 36 个显著增长到 113 个。

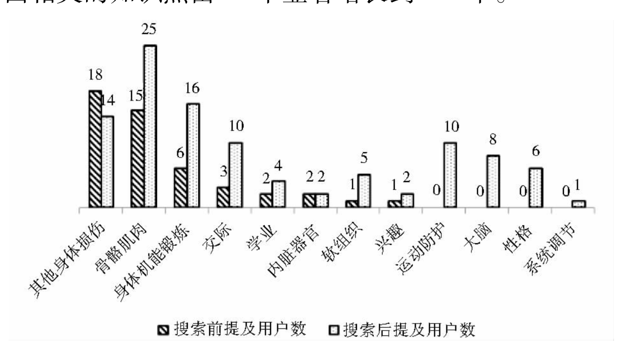


图 7　知识面的用户数量分布(单位:人)

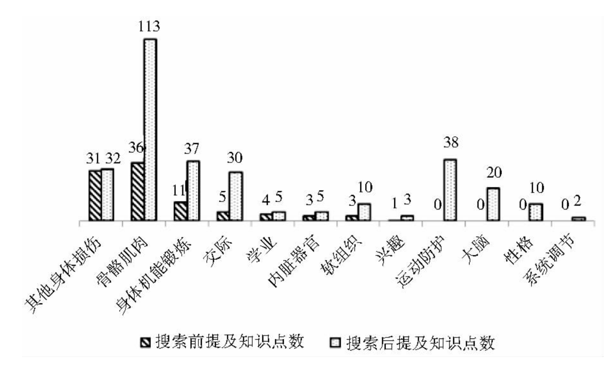


图 8　知识面包含知识点数量分布(单位:个)

综合来说,用户的知识在数量上的表现随着搜索的完成而变得更加全面和深入,在知识点数量、知识面数量、知识面广度和知识面深度上都有显著的提升。同时随着对任务理解的加深,伴随着一些专业度较高的知识面的产生,例如口罩任务中的材质、型号知识面,足球任务中的运动防护知识面等,用户对于搜索前较模糊的某些概念,也能在搜索后更清晰明确地表达。例如足球任务中,在搜索前用户集中于对其他身体损伤知识面的表达,是因为用户大都不明确具体损伤的身体部位,而在搜索之后,用户主要集中于对骨骼肌肉损伤的表达。

4.2 搜索前后用户知识质量的变化

通过分析发现,用户搜索前后在知识质量上存在显著差异。从表 5 中各项均值可知,用户在搜索前知识相关性、是否提出个人观点、知识的分析程度的平均得分分别为 0.84、0.13、0.39,可知大多数用户对于任务的知识较薄弱,从而导致他们很难提出自己的观点并进行深入的分析。随着用户搜索结束,用户知识的相关性和分析程度分别上升至 1.63 和 1.17,且过半数用户对任务相关话题产生了个人的观点。

表 5 用户搜索前后知识质量差异 W 检验

知识质量	搜索前(均值(方差))	搜索后(均值(方差))	比较
知识相关性	0.84(0.45)	1.63(0.43)	$p<0.0001^*$
用户观点	0.13(0.11)	0.66(0.23)	$p<0.0001^*$
知识分析程度	0.39(0.4)	1.17(0.68)	$p<0.0001^*$

注:$^*p<0.05$

图 9 是对用户搜索前后相关性、观点和分析程度变化的统计,由此可以更加清晰地展现出用户个体单位在搜索前后的变化。图中的横坐标各项代表搜索前后对应项得分的变化,例如"0 - 0"表示用户搜索前的某个指标得分为 0,搜索后对应指标上的得分仍然为 0。

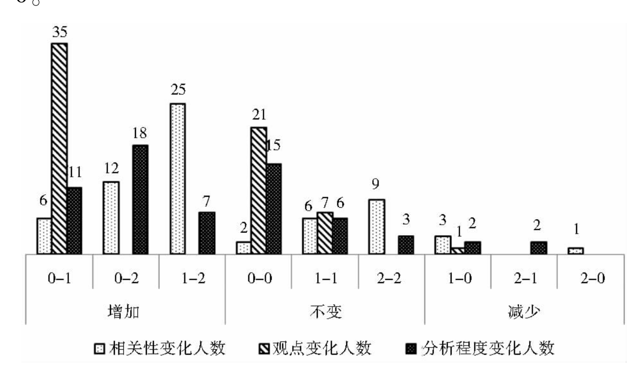


图 9 搜索前后用户知识质量变化统计(单位:人)

可以看出,在知识相关性上有所提升的用户共 43 人,包括"0 - 1"6 人、"0 - 2"12 人、"1 - 2"25 人;知识相关性上不发生变化的用户共 17 人,这其中在知识质量上真正没有提升的用户为"0 - 0"2 人和"1 - 1"6 人,至于"2 - 2"的 9 人,由于在搜索前回答的相关性已经达到一个比较良好的状态,从评分来看,没有继续上升的空间。

从用户是否产生个人观点来看,有过半数的用户(35 人)在搜索之前没有表达对于问题的个人观点,但在搜索之后形成了观点,即"0 - 1",但是仍有 21 人在搜索后也没有提出个人的观点,即"0 - 0",因此究竟是什么因素影响了用户观点的生成,观点从无到有的用户存在何种特质,也是本文后续需要关注的问题。

对于问题的分析程度,有 24 人搜索前后在分析程度上没有任何的变化,这其中包括 15 人搜索前后对于问题都几乎没有任何分析,即"0 - 0"。有共计 36 人在问题的分析程度上有所提升,包括"0 - 1"11 人、"0 - 2"18 人、"1 - 2"7 人。

综合来说,用户搜索后的知识质量较搜索前有了显著的提升,虽然本文也不排除少量用户在搜索之后可能会出现知识质量下降的情况,例如从有观点转变为无观点"1 - 0",从强分析程度转为弱分析程度"2 - 1"。这可能是由于搜索过程中用户花费了过多的精力和时间来理解和分析大量的信息,导致其在知识表达上产生倦怠的情绪。

5 讨论及结论

本文从知识数量与质量两个维度对用户搜索前后的知识水平进行评估并分析用户从搜索前到搜索后知识水平的变化。通过数据分析发现,用户搜索前后的知识水平存在显著的差异。具体来说,对于本文构建的知识评估体系中的知识数量(知识点数、知识面数、知识面广度、知识面深度)和知识质量(知识相关性、用户观点、知识分析程度)两个维度上所有的变量而言,用户搜索后较搜索前的知识状态都有显著的提升。根据 M. J. Wilson 和 M. L. Wilson[22] 的说法,这表明知识的这些评估维度对于用户真实的知识增长是灵敏的。本研究不同于 M. J. Wilson 和 M. L. Wilson[22] 研究之处在于,本文分析的是用户完成搜索后,在不参考任何网页和相关资料的情况下所撰写的答案,因此更加客观地展示了用户头脑中存在的知识状态,而不是在搜索过程中信息的搜集和整理。总体而言,本文设

计的知识评估体系对于测量用户的知识水平是有效的,同时从更加具体的层面上验证了已有研究所[11,26]提出的用户信息检索的过程往往伴随着知识的增长这一现象。

具体而言,本研究发现用户在知识数量上的表现随着搜索的完成而变得更加全面和深入,同时随着对任务理解的加深,伴随着一些专业度较高的知识面的产生(例如口罩任务中的材质、型号知识面,足球任务中的运动防护知识面等),对于搜索前较模糊的某些概念,用户也能在搜索后更清晰明确的表达。例如足球任务中,在搜索前用户集中于对其他身体损伤知识面的表达,是因为用户大都不明确具体损伤的身体部位,而在搜索之后,用户主要集中对于骨骼肌肉损伤的表达。在知识的相关性、分析程度及用户观点的提出方面,绝大多数的用户都在搜索后有所提升。

本文从多角度对用户搜索前后的知识水平进行评估,为研究者对信息搜索中用户的知识学习评估问题提供了可行的方案。本研究还发现搜索前后用户的知识水平存在一定的变化,且发现了这些变化的具体表现,在后续的研究中会进一步探索哪些因素会对用户搜索前后知识水平变化的程度产生影响,如任务类型、个人特征、交互行为或其他的情境因素是否会影响用户搜索前后知识水平的变化程度。

参考文献:

[1] RIEH S Y, COLLINS-THOMPSON K, HANSEN P, et al. Towards searching as a learning process[J]. Journal of information science, 2016, 42(1):19－34.

[2] GWIZDKA J, HANSEN P, HAUFF C, et al. Search as learning (SAL) workshop 2016[C]//Proceedings of International ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2016:1249－1250.

[3] RIEH S Y, GWIZDKA J, FREUND L, et al. Searching as learning: novel measures for information interaction research[J]. Proceedings of the American Society for Information Science & Technology, 2014, 51(1):1－4.

[4] MARCHIONINI G. Exploratory search: from finding to understanding[J]. Communications of the ACM, 2006, 49(4):41－46.

[5] 张云秋, 安文秀, 于双成. 探索式搜索中用户认知的实验研究[J]. 情报理论与实践, 2013, 36(6):73－77.

[6] ALLAN J, CROFT B, MOFFAT A, et al. Frontiers, challenges, and opportunities for information retrieval: report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne[J]. Comptes rendus des séances de la société de biologie et de ses filiales, 2012, 161(12):2355－2359.

[7] AGOSTI M, FUHR N, TOMS E, et al. Evaluation methodologies in information retrieval dagstuhl seminar 13441[J]. ACM SIGIR forum, 2014, 48(1):36－41.

[8] WHITE R W, ROTH RA. Exploratory search: beyond the queryresponse paradigm [C]//Proceeding of Synthesis lectures on information concepts retrieval and services. Williston: Morgan & Claypool Publishers, 2009: 1－98.

[9] 袁红, 李秋. 搜索任务和搜索能力对用户探索式搜索行为的影响研究[J]. 图书情报工作, 2015, 59(15):94－105.

[10] 张云秋, 安文秀, 冯佳. 探索式信息搜索行为研究[J]. 图书情报工作, 2012, 56(14):67－72.

[11] EICKHOFF C, TEEVAN J, WHITE R, et al. Lessons from the journey: a query log analysis of within-session learning[C]//Proceedings of ACM conference on Web search and data mining. New York: ACM, 2014:223－232.

[12] COLLINS-THOMPSON K, RIEH S Y, HAYNES C C, et al. Assessing learning outcomes in Web search: a comparison of tasks and query strategies[C]// Proceedings of ACM conference on human information interaction and retrieval. New York: ACM, 2016:163－172.

[13] LAZONDER A W, BIEMANS H J A, WOPEREIS I G J H. Differences between novice and experienced users in searching information on the World Wide Web[J]. Journal of the American society for information science & technology, 2000, 51(6):576－581.

[14] SAITO H, MIWA K. A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience[C]// Proceedings of the Second international conference on Web information systems engineering (WISE'01) IEEE Computer Society. Washington: IEEE Computer Society, 2001: 321－327.

[15] WHITE R W, DUMAIS S T, TEEVAN J. Characterizing the influence of domain expertise on Web search behavior[C]// Proceedings of the second ACM international conference on Web search and data mining. New York: ACM, 2009:132－141.

[16] KOROBILI S, MALLIARI A, ZAPOUNIDOU S. Factors that influence information-seeking behavior: the case of greek graduate students[J]. Journal of academic librarianship, 2011, 37(2):155－165.

[17] LIU C, ZHANG L, SONG X. Are self-assessment of search ability and performance reliable? [C]//Proceedings of the 79th ASIS&T annual meeting: creating knowledge, enhancing lives through information & technology. Maryland: American Society for Information Science Silver Springs, 2016:87.

[18] HERSH W R, ELLIOT D L, HICKAM D H, et al. Towards new measures of information retrieval evaluation [C]// International ACM SIGIR conference on research and development in information retrieval. New York: ACM, 1995:164－170.

[19] NELSON L, HELD C, PIROLLI P, et al. With a little help from my friends: examining the impact of social annotations in sense-making tasks[C]// Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM, 2009:1795 -1798.

[20] KAMMERER Y, NAIRN R, PIROLLI P, et al. Signpost from the masses: learning effects in an exploratory social tag search browser [C]// Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM, 2009:625 -634.

[21] WILLOUGHBY T, ANDERSON S A, WOOD E, et al. Fast searching for information on the Internet to use in a learning context: the impact of domain knowledge[J]. Computers & education, 2009, 52(3):640 -648.

[22] WILSON M J, WILSON M L. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries[J]. Journal of the American Society for Information Science & Technology, 2013, 64(2):291 -306.

[23] BLOOM B S, ENGELHART M D, FURST E J, et al. Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain[M]. London:Longmans,1956:58.

[24] LEE J. Does higher education foster critical and creative learners? An exploration of two universities in South Korea and the USA[J]. Higher education research & development, 2015, 34(1):1 -16.

[25] 刘畅, 赵瑜, 杨帆. 信息检索用户实验设计中时间限制和任务次序的影响研究[J]. 图书情报工作, 2015, 59(1):99 -105.

[26] LIU J, BELKIN N J, ZHANG X, et al. Examining users' knowledge change in the task completion process[J]. Information processing & management, 2013, 49(5):1058 -1074.

作者贡献说明：

宋筱璇:负责数据分析和论文撰写;

刘畅:负责研究设计,确定文章总体思路。

Assessment of User Knowledge Level Before and After Searching

Song Xiaoxuan　Liu Chang

Department of Information Management,Peking University,Beijing 100871

Abstract: [**Purpose/significance**] Information search extensively influence people's daily lives, and people tend to use search engine to find the information they need. However, the existing search system is better for supporting factual tasks search, and it is not satisfactory to support tasks such as searching in order to learn. In recent years, researchers have begun to pay attention to regarding search as learning, and try to evaluate the knowledge learning in search. [**Method/process**] For the purpose of mastering users' learning status in search thoroughly, in this paper, 32 students from Peking University were recruited to complete user experiment. User knowledge assessment method considering knowledge quantity and quality comprehensively was established in this paper. Based on this knowledge assessment method, user knowledge level in learning tasks before and after search was evaluated respectively, and the changes of knowledge level were investigated. [**Result/conclusion**] The research shows that users' performance on knowledge quantity dimension becomes more comprehensive and in-depth with the search completion, and has a significant improvement in the number of knowledge points, the number of knowledge facets, the breadth of knowledge facets and the depth of knowledge facets. At the same time, there are a number of relatively professional knowledge facets after searching. Some vague knowledge concept expressions before searching were expressed more clearly and explicitly after searching. In the dimension of knowledge quality, knowledge relevance, analysis and user opinion, the vast majority of users improve after searching. This paper provides a feasible way for researchers to evaluate user knowledge learning in information search from the perspective of multi-angle evaluation of knowledge status before and after searching.

Keywords: search as learning　knowledge assessment　knowledge change